

A Comparison of Unconstraining Methods to Improve Revenue Management Systems

Carrie Crystal^a
Mark Ferguson^{a*}
Jon Higbie^b
Rohit Kapoor^a

05/31/06

^aThe College of Management
Georgia Institute of Technology
800 West Peachtree Street
Atlanta, GA 30332-0520

^bManugistics, Inc
2839 Paces Ferry Road, SE, Suite 1000
Atlanta, GA 30339-5770

A Comparison of Unconstraining Methods to Improve Revenue Management Systems

Abstract

A successful revenue management system requires accurate demand forecasts for each customer segment. The forecasts are used to set booking limits for lower value customers to ensure an adequate supply for higher value customers. The very use of booking limits, however, constrains the historical demand data needed for an accurate forecast. Ignoring this interaction leads to substantial penalties in a firm's potential revenues. We review existing unconstraining methods and propose a new method that includes some attractive properties not found in the existing methods. We evaluate several of the common unconstraining methods against our proposed method by testing them on intentionally constrained simulated data. Results show our proposed method along with the Expectation Maximization (EM) method perform the best. We also test the revenue impact of our proposed method, EM, and “no unconstraining” on actual booking data from a hotel/casino. We show that performance varies with the initial starting protection limits and a lack of unconstraining leads to significant revenue losses.

Keywords: Revenue Management, Truncated Demand, Forecasting, Unconstrained Demand

1. Introduction

Revenue Management has been credited with improving revenues 3%-7% in the airline, hotel, and car rental industries (Cross, 1997). One of the core concepts behind revenue management is the reservation of a portion of capacity for higher value customers at a later date. The amount of capacity to reserve is typically determined through the calculation of booking limits, which place restrictions on the amount of capacity made available to a lower value segment of customers so as to reserve capacity for a higher value segment that may arrive in the future. Most booking limit calculations depend on the deduction of a demand distribution for each customer value segment from past demand data that occurred under similar circumstances and operating environments. In practice, however, true demand data is difficult to obtain as many firms are unable to record all demand request that arrive after a booking limit has been exceeded and capacity for that customer segment has been restricted.

To overcome this problem, “unconstraining” methods are used to extrapolate the true demand distribution parameters from truncated demand data collected over previous selling opportunities. Once a firm sells out of capacity for a given segment, the sales data for that segment represents truncated demand (equal to the booking limit) instead of true demand. While there is no perfect way to unconstrain sales data, Weatherford and Polt (2002) claim that, in the airline industry, switching from one common industry method to a better method increases revenues 0.5 to 1.0 percent. Since most firms using revenue management have low marginal costs, maximizing revenues translates into maximizing operating profits. Hence, unconstraining methods significantly impact revenues, and in turn, profits, and deserve closer research attention.

Despite the significant impact unconstraining has on the success of a revenue management application, this topic has received much less attention in the literature compared to the work on methods for setting and adjusting booking limits. This is surprising since the demand distribution

parameter estimates represent a primary input to most booking limit techniques, fundamentally linking the value of the former with the quality of the latter. A firm facing constrained sales data faces three choices: 1) leave the data constrained, 2) directly observe and record latent demand, or 3) statistically unconstrain the data after the fact.

If historical sales data is left constrained, true demand is underestimated, creating a spiral-down effect on total revenue where the firm's expected revenue decreases monotonically over time (Homem-de-Mello et al. 2006). Unfortunately, due in part to the absence of research and teaching in this area, this practice is common at firms using less sophisticated revenue management systems. In §5 we demonstrate how ignoring constrained data impacts revenue using actual booking data from a hotel/casino.

Direct observation involves the recording of latent (unsatisfied) demand. Many hotels for example, record both bookings (requests that are met) and turndowns (requests that are not met). Care must be taken however, as turndowns may be attributed to availability (denials) or price (regrets). The former is considered latent demand while the latter is not. To differentiate between the two, some hotel chains have invested in systems and training for their reservations agents to track turndowns, and rely on these direct observations to unconstrain their sales data. Unfortunately, there are many issues with using turndown data for unconstraining. These issues include multiple availability inquiries from the same customer, incorrect categorization of turndowns by reservations agents, and customer requests which arrive through a channel not controlled by the firm (Orkin 1998). The latter provides the largest hurdle for most industries.

Direct observations of demand are not an option for many industries because of their distribution channels. For example, traditionally most airline bookings have been made via travel agents using global distribution systems like Sabre and Worldspan, and no turndown information is collected on these

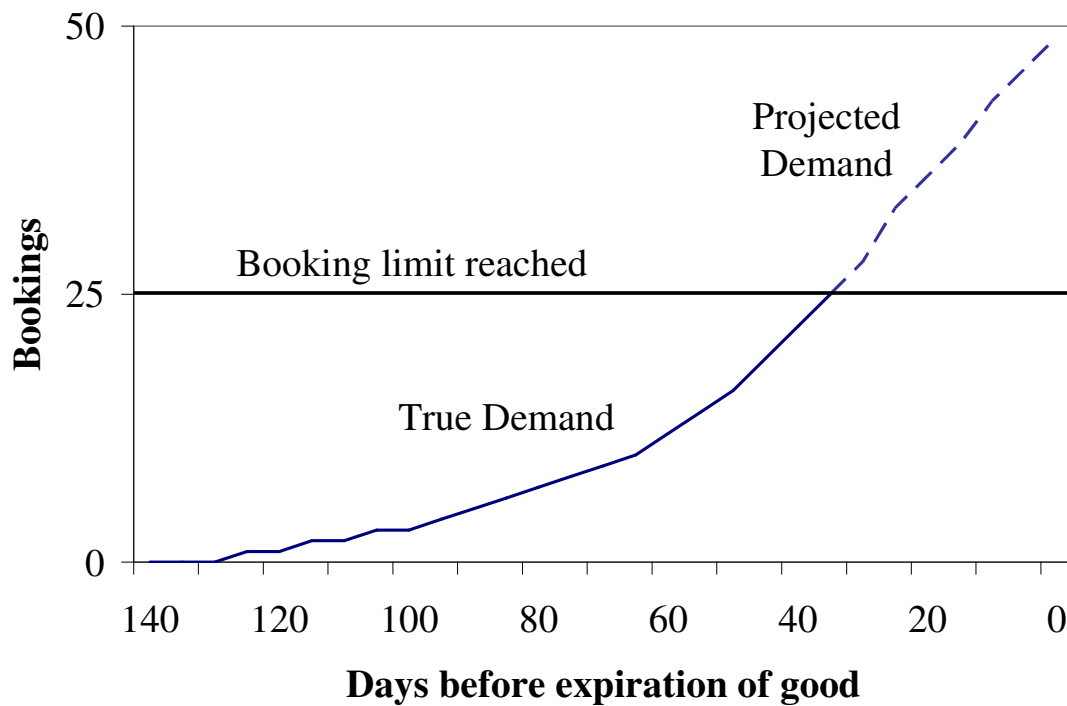
bookings. While airlines have recently been striving to increase their direct sales and improve their customer information, the percentage of total demand collected through firm-owned channels is still very small. On the other hand, hotels and casinos have historically taken the majority of their bookings through their own agents, either at the property itself or through a central reservations center. The advent of the Internet has compromised the quality of hotels' turndown data. While Internet direct sales is a growing channel for hotels, with some hotels taking up to 10% of their booking through this channel, most companies have yet to incorporate turndowns from their own web site into their total demand picture, and for good reason. Carroll and Siguaw (2003) point out that only 20% of hotel customers checking availability via the direct Internet channel actually book their rooms at the same site. Along with the growth in direct Internet sales, sales via third party web sites (such as Expedia and Travelocity) have grown at an even faster rate. Most third party web sites do not provide any turndown information. The net effect for hotels is an increasing proportion of bookings from channels with no turndown information, and as a result, hoteliers have an increasing interest in alternative unconstraining methods.

Statistical unconstraining covers a spectrum of optimization and heuristic techniques that rely only on observed bookings and a state indicator (open/closed). The purpose of this paper is to compare common statistical unconstraining methods to our proposed forecasting-based method. In addition, we apply the most accurate of these methods to actual hotel booking data. Previous studies on unconstraining methods tested a subset of the methods against simulated airline data.

Most traditional unconstraining methods follow a common methodology. Demand observed over time is categorized as constrained or unconstrained, then distribution parameter value estimates are adjusted based on the percentage of data that was constrained. In these unconstraining methods, a demand stream constrained twenty days before the end of the booking window is treated the same as a

demand stream constrained one day before. This methodology ignores an important aspect of the revenue management environment: firms often know the time periods when demand was constrained. Our proposed method takes advantage of this information and uses it when calculating the demand distribution parameter estimates. In addition, our method offers two other advantages over many of the alternative statistical methods: 1) It is based on a widely accepted statistical forecasting technique (double exponential smoothing) requiring minimal computations and 2) it is non-parametric, requiring no a-priori assumptions about the shape of the booking curve or the distribution of the final total demand.

Figure 1 - Example booking pace curve



We illustrate the key concept behind our proposed method through the example booking curve shown in Figure 1. Most traditional unconstraining methods only use the final observed demand and whether or not the demand was constrained. Our proposed method also uses the time demand was truncated (30 days before the the guest arrival in our example). Our method then uses a forecasting

technique (double exponential smoothing) to project total demand that would have been observed in the absence of any booking limits (the dashed part of the booking pace curve). Through a simulation experiment, we find our method outperforms most of the traditional statistical methods in estimating the demand distribution parameters of constrained data sets. Compared to the one method it does not always outperform, our method is simpler and works under conditions where the other method does not, such as when all historical data sets are constrained. Since there is no clear dominance by either method, we evaluate the impact on revenue the two methods provide using actual booking and revenue data from a leading hotel/casino.

The rest of the paper is organized as follows. In §2 we review the literature, in §3 we define the proposed method, in §4 we test the method against other common methods used in practice, and in §5 we test the two best performing methods on real hotel/casino data and measures their impact on the hotel's revenue. Finally, in §6 we offer conclusions and recommendations.

2. Literature Review

Weatherford and Bodily (1992) and McGill and van Ryzin (1999) provide general reviews of the broad range of literature in the revenue management field. Talluri and van Ryzin (2004b) provide an excellent overview of the current state-of-the-art in all aspects of revenue management systems. As these studies show, the primary research focus has been weighted towards the development of overbooking and booking limit techniques with little focus on unconstraining (also called uncensoring) sales data. We concentrate here on reviewing the unconstraining research.

Reliability engineers, biomedical scientists and econometricians have used unconstraining procedures for many years to compensate for the early termination of experiments. This parallels how revenue managers “terminate” demand for a particular customer segment through the use of booking limits. Relevant research in these fields include: (Cox, 1972; Kalbfleisch and Prentice, 1980; Lawless,

1982; Cox and Oakes, 1984; Schneider, 1986; Nelson, 1990; Liu and Makis, 1996). These methods rely heavily on the hazard rate function to determine the probability distribution of lifetime data. To our knowledge, van Ryzin and McGill (2000) provide the only use of this type method in a revenue management framework when they utilize a method based on demand lifetables. We include the lifetable method of uncensoring data in our comparison as described in Lawless (1982).

Weatherford and Polt (2002) and Zeni (2001) compare unconstraining methods using simulation and apply the best methods to an airline's reservation data to test the revenue impact of different methods. Six unconstraining methods are tested: three different averaging methods, a booking profile method, projection detruncation (PD), and expectation maximization (EM). The averaging methods are the simplest computationally and therefore are often used in practice. We compare our proposed method (DES) against the three best performing methods found in Weatherford and Polt (2002): the most accurate averaging method (referred to as Naïve 3 in Weatherford and Polt, abbreviated to AM in this paper), EM and PD. Both Weatherford and Polt (2002) and Zeni (2001) conclude the EM method outperforms the others and increases revenues by 2-12 % in full capacity situations. We also find the EM method outperforms PD and AM.

Of the three best methods that Weatherford and Polt (2002) and Zeni (2001) use, only EM is grounded in statistical theory. Dempster et al. (1977) prove the theory behind the EM method based on data from a univariate distribution. The EM method discussed by Dempster et al. (1977) is essentially the same as the tobit model used in econometrics (Maddala, 1983). McGill (1995) extends the EM method to a multivariate problem when demand for different classes (segments) of a good are correlated.

The PD method closely resembles the EM method, but takes a conditional median in place of a conditional mean. Additionally, the PD method allows users to change a weighting constant to obtain

more aggressive demand estimates. The tradeoffs include increased computation and increased risk of no solution convergence (Weatherford and Polt, 2002).

Liu et al. (2002) examine unconstraining demand data through the lens of the hotel industry and argue the EM method is unrealistic in application because of its computational intensity. The authors argue that parametric regression models take into account all relevant information and are computationally more feasible in real-world applications. They develop a parametric regression model which uses booking curve data, but requires knowledge of the shape of the demand distribution and other specifics of the demand constraints. This knowledge requirement restricts the general use of their model, as firms often do not know a priori the shape of the booking curve. Also, the authors do not provide comparisons between their proposed parametric method and the methods discussed in other papers. We do not include their method in our comparison because we do not assume a known, functional form for the booking curve. We do agree, however, with their criticism regarding the computational intensity of the EM method. Our proposed method is much easier to calculate but, unlike the parametric models, does not require knowledge about the shape of the booking curves.

To test the revenue impact of our new unconstraining method, we must set the protection levels effectively. To do so, we use the most common seat protection heuristic used in practice, EMSR-b (Belobaba 1989). McGill and van Ryzin (1999) give an explanation of the EMSR-b method along with a review of the booking limits problem in general.

3. Proposed Unconstraining Method

Our proposed method uses Double Exponential Smoothing (DES) or “Holt’s Method” to forecast the constrained values of a given data set. DES uses two smoothing constants: one for smoothing the base component of the demand pattern and a second for smoothing the trend component. Armstrong (2001) provides a good review of this method. Below, we describe how it may be used to solve the

unconstraining problem where demand is constrained only once, prior to the final period. In the appendix we provide an example of how this method can be used to solve the unconstraining problem where demand is constrained multiple times due to re-opening closed booking classes

Let t represent the time periods between I , the period that reservations are initially accepted, and B , the period where demand reaches the booking limit (time is counted backwards from I days before arrival until B days before arrival). That is $t \in [I, I-1, \dots, B], I \geq B$. After period B , demand continues to occur but is unobserved. If demand is never constrained then $B = 0$. Thus, “demand seen” equates to the cumulative demand observed from periods I to B and is always less than or equal to true demand. From Figure 1, I corresponds to period 140 and B corresponds to period 30, after which demand is unobserved. Now define:

A_t = Actual cumulative demand in period t

F_t = The exponentially smoothed base component for period t

T_t = The exponentially smoothed trend component for period t

FIT_t = The forecast of cumulative demand including trend for period t

α = Base smoothing constant

β = Trend smoothing constant

The forecast for the upcoming period t is

$$FIT_t = F_t + T_t \quad (1)$$

where

$$F_t = FIT_{t+1} + \alpha(A_{t+1} - FIT_{t+1}) \quad \text{and} \quad (2)$$

$$T_t = T_{t+1} + \beta(F_t - FIT_{t+1}). \quad (3)$$

The smoothing constants, α and β , are decision variables. For each constrained demand instance, we use a non-linear optimization routine to select the α and β values that minimize the sum of the squares of the forecast error:

$$\min_{\alpha, \beta} \sum_{t=I}^B (A_t - FIT_t)^2 \quad (4)$$

For the initial values, F_I and T_I , we use the actual demand in period I as our estimate for the base component and the average trend over the available observed cumulative demand as our estimate for the initial slope component. Since the problem is not jointly convex in α and β , a non-linear search algorithm such as tabu search or simulated annealing is needed to find the global minimizers. The forecasting model is then used to project the cumulative demand over the periods in the data set where demand is constrained, i.e. over periods $B-1$ to 0 . It does so by taking the last forecast where it was possible to update with observed demand, $FIT_B = F_B + T_B$, and projecting the final B periods where demand is constrained: $FIT_0 = F_B + BT_B$.

We demonstrate our method using the booking curve shown in Figure 1. First, initial estimates of F_{140} and T_{140} are calculated using the cumulative demand in period 140 and the average trend between period 140 and period 30 (the last period in which we observe unconstrained demand). If $A_{140} = 1$ and $A_{30} = 25$, then $F_{140} = 1$ and $T_{140} = (25-1)/30 = 0.8$. Next, we choose an initial value of 0.1 for both the α and β smoothing coefficients and perform one-period ahead forecasting to find FIT_t for periods $t = 140 - 30$ (the initial starting values chosen for the smoothing coefficients are not important as long as we use a search algorithm that does not get stuck at local minimums to find the smoothing coefficients that best fit the data available). We then choose the α and β smoothing coefficients that minimize

$\sum_{t=30}^{140} (A_t - FIT_t)^2$. Let $\bar{\alpha}$ and $\bar{\beta}$ represent the smoothing coefficients that result from this minimization

search. These smoothing coefficients are used to recalculate FIT_t ; call the new forecasts $\overline{FIT}_t = \overline{F}_t + \overline{T}_t$, for periods $t = 140 - 30$. Because we know the actual demands that occurred during days 140 to 30, we calculate \overline{FIT}_{140} to \overline{FIT}_{30} using the recursive methods of (1), (2) and (3) for each day until we reach $\overline{FIT}_{30} = \overline{F}_{30} + \overline{T}_{30}$, after which we no longer know true demand and can no longer update our forecast. The remaining 30 periods ($t = 29 - 0$) are when demand is constrained in our example. Our objective is to determine the final cumulative demand (if demand was not constrained) at the terminating period $t = 0$. We estimate this value using $\overline{FIT}_0 = \overline{F}_{30} + 30\overline{T}_{30}$.

The cumulative demand over the observed and projected components of the booking curve (\overline{FIT}_0) is then used as a single point estimate of true cumulative demand for a particular selling occurrence (i.e. a given Thursday night stay for a given fare class at a hotel). Call this individual point estimate for the i th booking curve X_i . We repeat this procedure over each constrained booking curve in a given data set (i.e. all Thursday night stays for a given fare class at a hotel). Thus, if there are n historical booking curves in the data set, we end up with a set of point estimates (X_1, X_2, \dots, X_n) . The final demand distribution parameters (mean μ and variance σ^2) are then estimated using this set of point estimates by:

$$\bar{\mu} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{and} \quad \bar{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{\mu})^2}{n} . \quad (5)$$

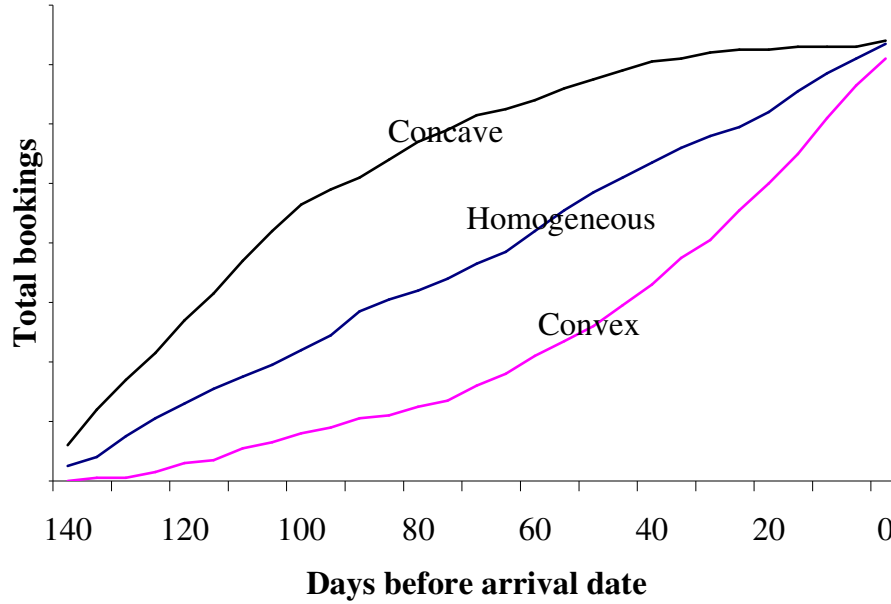
The basic model of DES described above is a very general method for forecasting demand and, as presented, does not account for seasonality, intermittent demand, and other specifics that might be relevant in application. However, DES can be easily adjusted to incorporate seasonality (Armstrong, 2001). In Section 4.2, we provide an alternative formulation that can be used when total demand is small and intermittent.

4. Comparison of Unconstraining Methods

In this section we compare four of the most common statistical unconstraining methods that have previously appeared in the literature with our proposed DES method. To compare the performance of the different methods, we simulate booking curves representing true demand and then impose booking limits to create constrained data. We apply five different unconstraining methods to the constrained data sets and compare the estimated demand parameters against the true parameters. We define the best unconstraining method as the method that estimates the demand distribution parameters closest to the true parameters.

To compare the performances of the chosen methods, we first simulate booking curves and set booking limits to constrain the demand data. To test each unconstraining method against a broad range of demand scenarios, we simulate three data sets with 100 booking curves each and 140 days in each booking curve. The three data sets represent three common shapes of booking curves: convex, homogeneous and concave (Liu et al., 2002) as shown in Figure 2. The 100 booking curves represent 100 historical demand records (for each shape curve) a hotel or airline may use to predict future demand. For example, a hotel may keep demand data from its last 100 Thursday night stays in order to estimate demand for future Thursday night stays. Since most hotels and airlines observe the majority of their reservations within 140 days before the day of arrival or departure, we simulate 140 days of daily demand arrivals for each booking curve, resulting in 100 individual booking curves of 140 days each, or 14,000 individual data points. For each booking curve shape, we look at demand observed for all 100 booking curves simultaneously (some where the total demand was not constrained and others where total demand exceeded the booking limits) and use each unconstraining method to estimate the true demand distribution parameters.

Figure 2 - Concave, Homogeneous and Convex booking curves



To construct the booking curves, we assume arrivals on a given day are randomly drawn from a Poisson distribution. This assumption is common in the literature and matches closely with actual data from the hotel and airline industries (Rothstein, 1974; Bitran and Mondschein, 1995; Bitran and Gilbert, 1996; Badinelli, 2000; Liu et al., 2002). For the homogeneous booking curve, we maintain a constant mean arrival rate over all 140 days. For the convex (concave) booking curves, we increment the mean arrival rate from low to high (high to low) respectively, so the expected total demand over the 140 day period is approximately equal for all three curves (average demand for concave, homogeneous and convex curves is 700, 698, and 696, respectively).

After we create the demand curves, we calculate booking limits. A user sees the minimum of true demand and the booking limit. A simple example is shown in Table 1. If daily demand arrivals are Poisson and the demands on different days are independent, then total demand is again Poisson. Because the mean of the Poisson-distributed total demand is sufficiently large, the distribution of total demand is approximately Normal. Thus, we calculate an expected average (μ) and standard deviation

(σ) of the total demand and generate five sets of booking limits representing various ranges of constraint levels. For example, a 20% constraining level means that, on average, 20% of the data sets have their total demand constrained by the booking limit. To find the booking limits at these various levels, we use the z -score from a standard Normal distribution corresponding to the 20%, 40%, 60%, 80%, and 98% constrained levels, where z represents the number of standard deviations above or below zero for a standard Normal distribution. Thus, to find the z -score corresponding to 98%, we find the point where the area under the standard Normal curve equals 0.98, or $z = 2.05$. We then set our corresponding booking limits using: $\mu + z * \sigma$. (See the appendix for more explanation on how we constructed these curves and set our booking limits to constrain them)

Table 1 - Example of true demand vs. demand observed

True Demand	100	110	91	95	103
Booking Limit	98	105	103	99	102
Demand Seen	98*	105*	91	95	102*

* indicates constrained demand, also called a closed segment

We test the five unconstraining methods across the three booking curve shapes (homogeneous, convex, and concave) for each of the five constraining levels to test how each method performs under varied conditions. We chose the unconstraining methods to test from previous research; the first three methods are the best performing methods from Weatherford and Polt's (2002) comparison. These methods include: 1) an averaging method (AM), called Naïve #3 by Weatherford and Polt, 2) Projection Detruncation (PD) and 3) Expectation Maximization (EM). Medical and reliability engineering researchers commonly use lifetables (LT), the fourth method. Additionally, van Ryzin and McGill (2000) use lifetables in a revenue management context. We provide a short description of each of these methods in the appendix. The fifth method is DES, which was described in §3.

4.1 Results of Comparison

Overall, the EM and DES methods outperform the other three unconstraining methods. Table 2 compares the percentage error for each unconstraining method versus the actual mean of the demand distribution (the percentage errors were similar but slightly larger for the estimated variances). DES outperforms all methods for the homogeneous and convex data sets as its error remains less than 0.5% for all levels of constraining, compared to a maximum 5% error for the other methods. In the concave data set however, EM outperforms DES on average. Table 2 summarizes the results of the comparison and Figure 3 graphically summarizes the mean absolute error over all three curves.

Previous comparisons (Zeni, 2001; Weatherford and Polt, 2002) show EM outperforming PD; we confirm this result. Aside from the accuracy issues, PD has two disadvantages compared to EM: it takes more iterations to converge than EM and it requires the choice of a weighting parameter, τ , creating an opportunity for varying results. A $\tau < 0.5$ can lead to better results, but increases both the time to convergence and the chance for no convergence. As seen in Table 2, AM performs the worst, although in some instances the difference between it and the others is small. Thus, under some conditions, practitioners may be justified in forgoing the increased accuracy of the other methods for the computational simplicity of the AM. Note the methodology of the AM (see appendix for details) ensures it will always underestimate the true distribution mean.

The LT method of unconstraining data produces estimates with errors very close to zero and even outperforms the DES and EM methods in a few of the concave cases. However, this method requires many computations and a large quantity of historical demand data. In a dynamic environment such as the travel industry, customer demand data changes quickly due to changes in the economic climate, broader market supply-demand-price relationships, and customer preference. Because of this, sufficient historical demand is often not available for the LT method to produce effective results.

Table 2 - Percentage error between calculated and actual mean for each unconstraining method

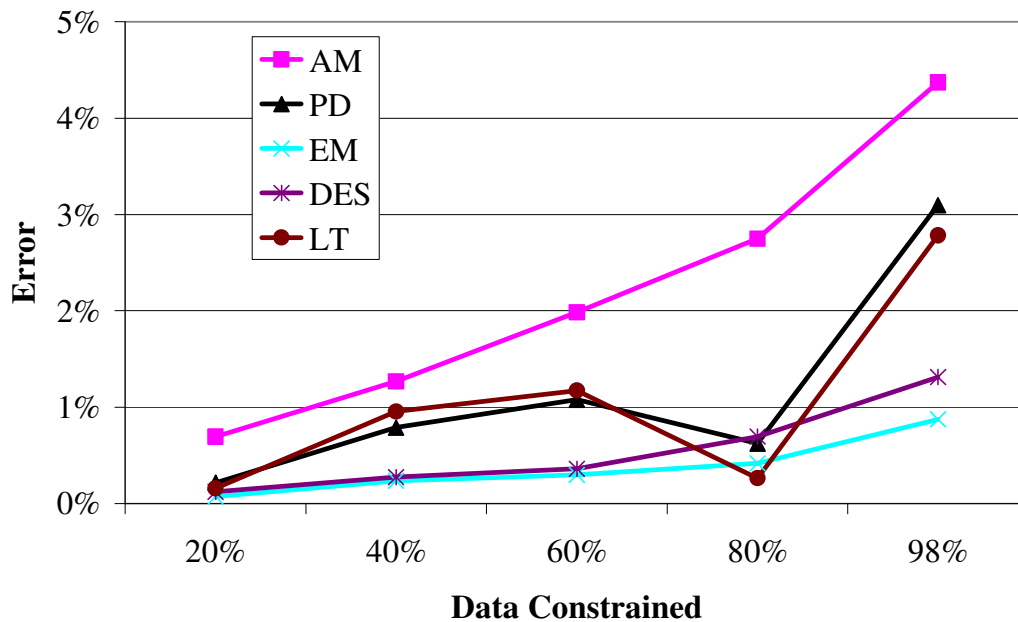
Booking Curve	Method	Percent of Data Sets Constrained				
		20%	40%	60%	80%	98%
Homogeneous	AM	-0.57%	-1.29%	-2.15%	-2.87%	-4.30%
	PD	0.23%	0.43%	0.56%	-0.53%	-2.99%
	EM	-0.06%	-0.23%	-0.56%	-0.22%	-0.58%
	LT	-0.17%	-1.31%	-1.53%	0.20%	0.43%
	DES	0.00%	0.00%	0.00%	0.00%	-0.14%
Convex	AM	-0.72%	-1.01%	-1.87%	-2.73%	-4.17%
	PD	0.26%	1.29%	0.98%	-0.91%	-2.99%
	EM	-0.08%	0.39%	-0.25%	-0.85%	-1.10%
	LT	0.13%	0.76%	0.89%	0.55%	5.72%
	DES	0.00%	0.14%	0.00%	-0.14%	-0.29%
Concave	AM	-0.71%	-1.43%	-1.86%	-2.57%	-4.57%
	PD	0.24%	0.73%	1.78%	-0.35%	-3.24%
	EM	-0.08%	-0.09%	0.09%	-0.19%	-0.93%
	LT	-0.21%	-0.83%	1.14%	-0.11%	2.29%
	DES	0.29%	0.71%	1.00%	1.86%	3.43%
Mean absolute error over all 3 Booking Curves	AM	0.67%	1.24%	1.96%	2.72%	4.35%
	PD	0.24%	0.82%	1.11%	0.60%	3.07%
	EM	0.07%	0.24%	0.30%	0.42%	0.87%
	LT	0.17%	0.97%	1.19%	0.29%	2.81%
	DES	0.10%	0.28%	0.33%	0.67%	1.29%

For the homogeneous data pattern, the DES method has negligible error across the range of constrained data sets due to the high predictability when arrival rates are constant over a given time period. For this data pattern, the DES method estimates the distribution mean up to 4% closer to the true mean than the next closest method. Similarly, for the convex data set, the DES method provides the most accurate estimate over all of the constraining conditions.

The DES method does not perform as well on the concave data set, although it still performs within a 1% error until demand is constrained in over 80% of the observations. DES underperforms on this demand pattern because booking segments close farther away from the arrival date for the concave booking curve, so many more data points must be estimated compared to the convex or homogeneous demand patterns. Here, the trend component of DES affects its accuracy as high demand occurring early

in the booking curve is projected to continue once the booking limit has been met. For this booking curve shape, a forecasting method with a trend that is dampened over time may perform better. In practice however, an inaccuracy in unconstraining demand following a concave demand pattern is not a great concern. Airlines often offer cheaper fares to customers booking at least three weeks in advance. Because of this, and other similar restrictions, the lowest valued segment is often forced to follow the concave demand pattern. Due to the fundamental concepts behind revenue management, errors in estimating the true demand for the lowest valued segments are typically less costly than are errors in estimating the demand for higher valued segments.

Figure 3 - Average absolute error from true demand for each unconstraining method

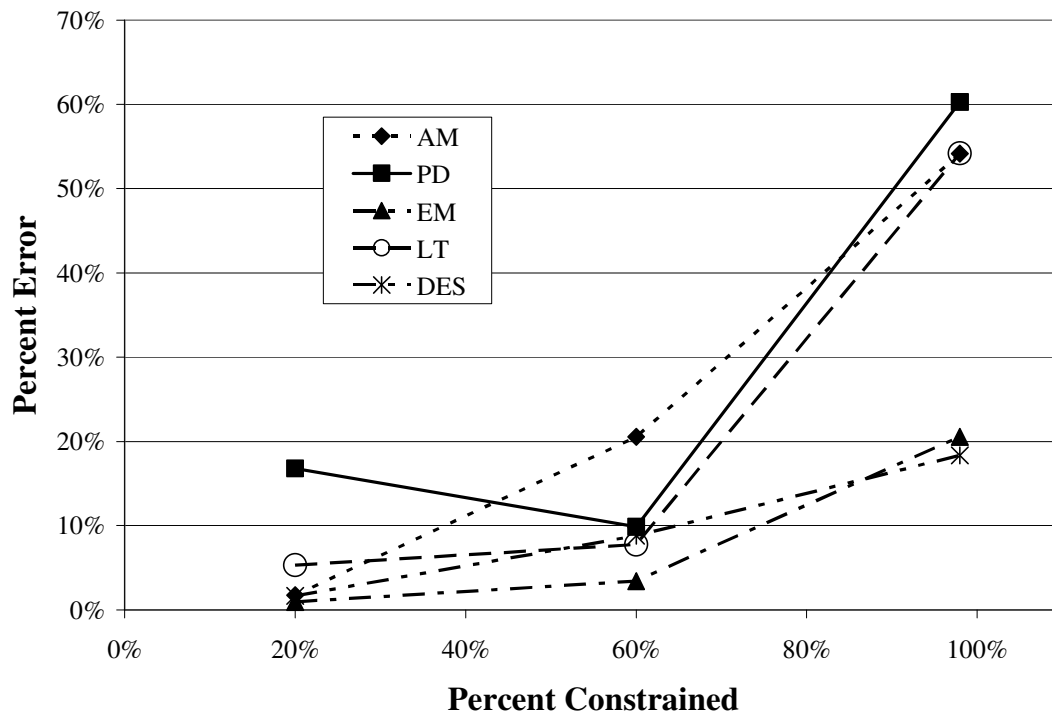


4.2 Performance with Smaller Demand

The first set of results (Table 2) compares unconstraining methods when total demand averaged 698 units. However, in many applications, total demand is much smaller than 698, so we run a similar

experiment with an average total demand of 19. We call this the “Small Demand” data set. Just as before, we ran simulations on homogeneous, concave, and convex booking curve shapes, with 100 trials of 140 days each for each shape. The DES and EM methods provide the most accurate results across a range of constrained data, and the AM provides the least accurate results. Due to difficulties of predicting data with intermittent demand (many periods with zero demand), the small demand data has less accurate results than the large demand data. This observation is consistent with previous studies; goods with intermittent demand are difficult to forecast and require specialized forecasting tools for the most effective results (Altay and Litteral, 2005).

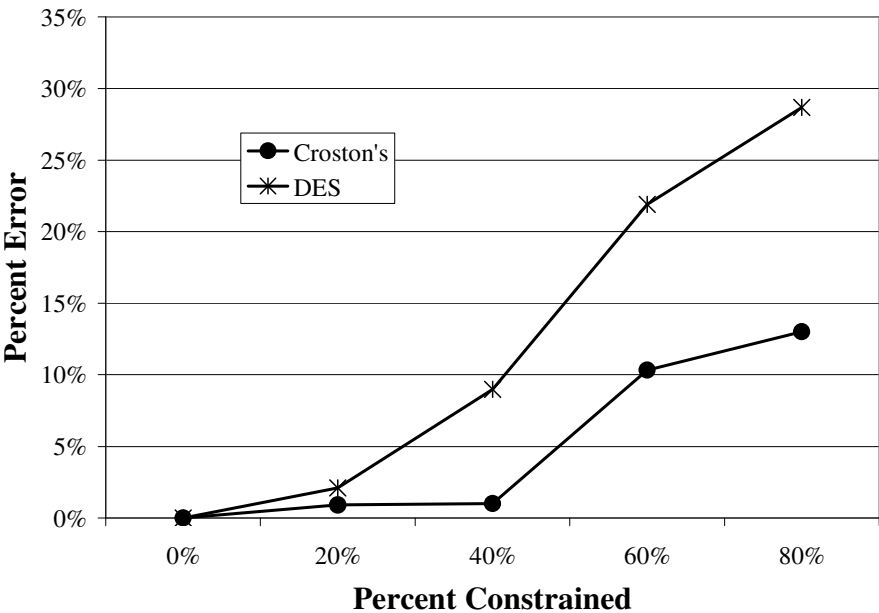
Figure 4 - Unconstraining error with small demand data set



Because DES has significantly higher error with the small demand data set than with the large demand data set, we sought an alternative formulation. Croston’s forecasting method (Croston, 1972) is a simple exponential smoothing method designed to accommodate small or intermittent demand. This method forecasts the size of the non-zero demands and inter-arrival time between non-zero demands.

(See the appendix for explanation of this method) In a simple simulation over 60 trials and 60 days with a total demand of 12 (based on the smallest observed demand segment of our partner hotel), Croston’s method outperforms DES across the range of constraints, as shown in Figure 5. These results provide evidence that Croston’s method may be superior to DES for unconstraining intermittent demand. When the percentage of days with zero demand exceeds 10% of the total number of days in the booking curve, Croston’s method begins to outperform DES.

Figure 5 - Unconstraining error with small demand data set – DES vs. Croston’s method



5. Revenue Impact Using Industry Data

In this section we compare the potential revenue impact of a major hotel/casino using DES, EM, and no unconstraining. Since unconstraining methods only provide estimated parameters for the demand distribution, we use the EMSR-b (Expected Marginal Seat Revenue) algorithm (Belobaba, 1989) - a

widely accepted method for setting booking limits for a basic revenue management system - to translate the demand distribution parameters (and corresponding room rates) into booking limits. We applied the booking limits to booking data from a hotel/casino to calculate the total revenue impact. Thus, we compare the revenue convergence using EM, DES, and “ignoring unconstraining” based on actual (but normalized) booking and revenue data from a major hotel/casino. While the examples presented in this section are very useful for illustrating the effectiveness of the methods, they cannot lead to conclusions about industry performance. Such conclusions can only be drawn from trials in practice.

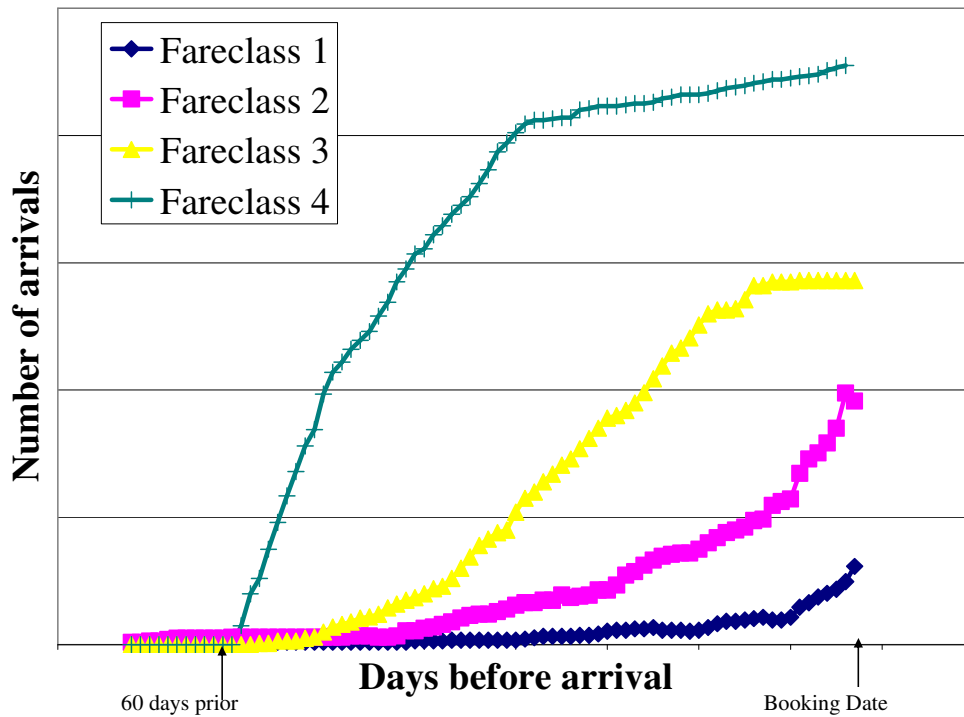
5.1 Demand Data

We use actual hotel/casino booking data to test how unconstraining impacts revenue. We use booking curve data for 12 consecutive Friday night stays, unconstrained using direct observation of turndowns. Extra care was taken to ensure that all demand was captured for this data set including demand that occurred after booking limits were met. Because of the increased cost involved in such careful data collection, we limited the data collection period to 12 weeks and used bootstrapping to create 1000 booking curve samples from the initial data. Hotel reservations vary greatly by day of the week, depending on the type of hotel. For this hotel, weekends are the most popular, and therefore, have the highest constraining rate. In order to control for differences in demand between different days of the week, we focus on Friday night stays during the 12-week period. Within any Friday night’s booking data, this hotel/casino has many different customer segments, with some customer segments so valuable they are rarely constrained (revenue per night from the highest fare customer can be 12 times the revenue from the lowest fare customer); therefore we focused our unconstraining efforts on the most popular four segments that are constrained.

We base many of our simulation choices on the hotel/casino data characteristics. In order to understand these characteristics, we plotted the unconstrained demand data for the four customer

segments for 12 consecutive Friday night stays. The shape of a given customer segment arrival rate was consistent throughout the 12 week period, although each of the 4 customer segments had a distinctly different booking curve. We present one such diagram in Figure 6 to show the variety of booking curve shapes. Based on the similar arrival rates throughout the 12 week period, we utilized bootstrapping to create a sufficient number of different demand realizations for our simulation study.

Figure 6 - Cumulative hotel/casino reservations for four separate fare classes



We created bootstrapped samples as follows. The 12 Friday night stay booking curves showed the slope of the curve changed dramatically at different intervals before arrival. Based on these slope changes, we created multiple intervals within the 60 day window which had similar arrival rates. Picking randomly (with replacement) from 12 weeks' worth of Friday night booking patterns within a similar arrival rate interval, we used the bootstrap method with replacement to create 1000 different 60 day booking curves for each of the four customer segments.

The hotel would temporarily close the lower valued segments midway through the 60 day booking curve, thus we simulated this practice of opening and closing the classes multiple times. We closed a fare class (constrained demand) midway through the booking curve, reopened the fare class, and then closed it again before the actual day of arrival. We initially closed the fare class if cumulative demand reached the first booking limit within 30 days of accepting bookings. We reopened the fare classes with 30 days left in the booking period and closed the fare class again if demand reached the second booking limit. We set protection levels so that 50% and then 75% of a given data set would be constrained to test our methodology against different constraining levels. To set protection levels so that 50% of a given data set would be constrained, we calculated the first booking limit as the average cumulative demand after 30 days of accepting bookings; the second booking limit was the average cumulative demand over the 60 day period. Similarly, for 75% constrained data, we used the average and standard deviations at 30 and 60 days to find the z-score corresponding to 25% unconstrained data ($\mu + z \cdot \sigma$). We did this for each of the 4 different segments in all 1000 replications.

Using both DES and EM, we unconstrained the data sets and compared the distribution parameter estimates for each method against the true parameter values. Both methods performed well, with average errors listed in Table 3. Just as in previous trials, all methods performed better with less constrained data. The methods better predict mean values than standard deviations. Over 1000 instances, both methods predict the mean within 5% of the true mean, showing the methods performed well even when demand was constrained multiple times in a booking curve.

Table 3 – Error comparison between EM and DES with interrupted arrivals and actual data

		50% of Data Sets Constrained	75% of Data Sets Constrained
Mean	EM	0.84%	1.72%
	DES	0.84%	4.87%
Standard Deviation	EM	7.38%	14.72%
	DES	7.65%	23.86%

While providing a useful comparison, such a simple test as the one above does not, however, capture the true impact the use of an unconstraining method has in practice. Three main issues are yet to be addressed: 1) in practice a hotel will never have 1000 historical booking curves available to estimate a demand distribution during a period of stationary demand; 2) protection levels evolve over time (the study above measures does not capture the transient nature of demand information updating or learning); and 3) there are no previous studies linking the accuracy of an unconstraining method to the revenue impact of the user. We attempt to address these issues in the following study.

In the next study, we test the impact each unconstraining method has on the hotel's total expected revenue. Total expected revenue from a revenue management system is the ultimate indicator of a system's success, but unconstraining methods only provide estimates for the demand distribution parameters. Thus, we borrow van Ryzin and McGill's (2000) general methodology for translating protection levels into revenue. To test convergence and robustness, we start with purposefully high and low protection levels, similarly to van Ryzin and McGill (2000). Note our switch to protection levels rather than booking limits for this study. Protection levels are the opposite of booking limits, i.e. how many units of capacity at a given class to protect for higher fare classes. To illustrate, consider a hotel

with two fare classes, high and low, and a total capacity of 100 rooms. If a booking limit for the low fare rooms is set at 60, the corresponding protection level for the high fare rooms is $100 - 60 = 40$ rooms protected for high fare customers.

Protection levels must be set at some estimated level for initial product offerings and for existing products when there has been a fundamental shift in the underlying demand distributions. As more demand is observed over time, the firm adjusts protection levels accordingly to increase total revenue. The convergence rate to the optimal protection levels depends on both the starting levels chosen and the unconstraining method used. Thus, we test protection level convergence and total revenue convergence using the two best performing unconstraining methods (EM and DES) with two different starting protection levels – low and high. To underscore the importance of unconstraining, we included data with no unconstraining, labeled (Spiral) for the decreasing revenue named the “spiral-down effect” (Homem-de-Mello, et al 2006) which occurs when data is not unconstrained.

5.2 Setting Protection Levels: the EMSR-b Method

For setting the protection levels for the hotel rooms, we use a variation of the EMSR (Expected Marginal Seat Revenue) heuristic (Belobaba 1989), called EMSR-b. This is the most common heuristic used in practice for setting protection levels. The EMSR-b method does not produce optimal protection levels under all real world conditions, but is representative of a basic revenue management system and is sufficient for comparing unconstraining methods. EMSR-b works as follows: Given the estimates of the means, $\hat{\mu}_i$ and standard deviations, $\hat{\sigma}_i$ for each customer value segment i , the EMSR-b heuristic sets protection level θ_i so that $f_{i+1} = \bar{f}_i P(\bar{X}_i > \theta_i)$, where \bar{X}_i is a normal random variable with mean

$\sum_{j=1}^i \hat{\mu}_j$ and variance $\sum_{j=1}^i \hat{\sigma}_j^2$, f_i is the fare for customer value segment i and \bar{f}_i is weighted average revenue

from classes 1, ..., i , given by $\bar{f}_i = \sum_{j=1}^i f_j \hat{\mu}_j / \sum_{j=1}^i \hat{\mu}_j$.

In simpler terms, this rule performs a marginal analysis on the benefits of holding capacity for a higher valued customer versus the cost of turning away the next lower valued customer. To demonstrate the EMSR-b method, consider the data given in Table 4 representing a hotel with four fare classes. The nested protection levels calculated using the EMSR-b method are given in the far right column (note that fractional values are rounded up). Thus, for this example, 49 rooms should be reserved for fare class 1 customers (the highest paying), 125 rooms should be reserved for fare class 1 and 2 customers, 257 rooms should be reserved for fare class 1, 2, and 3 customers, and any remaining rooms can be sold to the lowest fare class customers (class 4).

Table 4 – Example of protection levels set using EMSR-b heuristic

Class	Fare	Mean	Variance	Protection level θ_i
1	250	50	50	49
2	150	75	75	125
3	100	125	125	257
4	50	500	500	Capacity

The example above shows how demand distributions can be converted to recommended protection levels for a revenue management system. However, revenues can only be optimized if the true demand distribution parameters are known; hence the need for a good unconstraining method.

5.3 Simulation

We test the revenue impact of the unconstraining methods by applying protection levels (based on the EMSR-b method using the distribution parameter estimates from the unconstraining method) to the industry data described in §5.1. As in van Ryzin and McGill (2000), we assume nested fare-class

allocations, low fare classes book strictly before high fare classes, demand for each fare class is independent, and there are no cancellations or no shows. We define notation in the following manner: X_i = demand for fare class i ; I_i = inventory (hotel rooms) available for fare class i ; n = number of fare classes; and R_i = revenue earned from fare class i . We can now calculate total revenue using:

$$\text{Total Revenue} = \sum_{i=1}^n R_i \quad \text{where} \quad (6)$$

$$I_i = \begin{cases} \text{Capacity} - \theta_{i-1} & \text{for } i = n \\ \text{Capacity} - \theta_{i-1} - \sum_{j=i+1}^n \min(I_j, X_j) & \text{for } 1 < i < n \\ \text{Capacity} - \sum_{j=i+1}^n \min(I_j, X_j) & \text{for } i = 1 \end{cases} \quad \text{and} \quad (7)$$

$$R_i = f_i^* \min [X_i, I_i]. \quad (8)$$

(See the appendix for an example of this revenue calculation with a particular demand realization)

Normalizing the fare class data from our hotel/casino on a scale from \$1 - \$100, the per-night expected revenues for the four customer segments are: \$25, \$35, \$62, \$100. These expected revenues are based on the total amount a customer in that segment is expected to spend at the hotel/casino per night, including the revenue from the room rate, food, beverages, shows, and casino. The hotel/casino tracks customer spending by issuing frequent stay cards which record each time the customer makes a transaction.

Because estimates for the parameter values of the demand distribution and the corresponding protection levels, evolve over time, we simulate this evolution in our study. First, we split each of the 4 sets of 1000 booking curves described in §5.1 into 10 sets of 100 booking curves. Working with the first set of 100 booking curves, we estimate initial protection levels for each customer segment (two initial starting protection levels are used for each segment, one lower and the other higher than the optimal protection levels). These initial protection levels are rough estimates of demand and equate to

rough demand estimates when a firm offers a new product. Hence, with high protection levels, we set the protection levels to approximately equal average demand plus one standard deviation for the three highest fare classes. Similarly, with low protection levels, we set the protection levels to approximately equal average demand minus one standard deviation for the three highest fare classes.

We then calculate the revenue the hotel/casino would have received if they used these initial protection levels for each customer segment over all 100 booking curves in the set. The first data point in Figures 7 and 8 is the percentage difference in revenues the hotel/casino would have received using these initial protection levels versus if they had used optimal protection levels calculated with the true demand distribution parameters. Next, we applied the initial protection levels to the first 10 booking curves (booking curves 1-10 out of the 100 in the set). Thus, some segments of the first 10 booking curves were constrained by the calculated protection levels. We apply each of the unconstraining methods to this group of 10 constrained booking curves and calculated new protection levels for the next group of 10 booking curves (booking curves 11-20 out of the 100 in the set). Based on these new protection levels, we calculated the revenue generated if these protection levels were used on all 100 booking curves in the set. We continue this procedure, unconstraining the demand data and readjusting the protection levels every 10 booking curves. This procedure simulates a hotel manager watching demand for 10 consecutive Fridays, then adjusting his protection levels for the next 10 Fridays, and continuing this procedure for a total of 100 consecutive Fridays.

For robustness, we applied this methodology to 10 sets of 100 booking curves to calculate a standard error of our estimates. Figures 7 and 8 represent average results over the 10 sets along with upper and lower limits corresponding to a 95% confidence interval. All of the methods (EM, DES, Spiral) use the same data sets (the simulations are coupled)

We compare revenues for each of the unconstraining methods to optimal revenue. We calculate the optimal revenue by finding the mean and standard deviation of each set of 1000 booking curves (one set of 1000 for each of four customer segments). Since the initial data was unconstrained, we know the true demand for every booking curve, and hence know the true mean and standard deviation parameter estimates. We apply EMSR-b using these parameters to find protection levels, then using nested protection levels, apply (6 - 8) to calculate total revenue.

Figure 7 – Revenue achieved using high protection limits for EM, DES, and no unconstraining

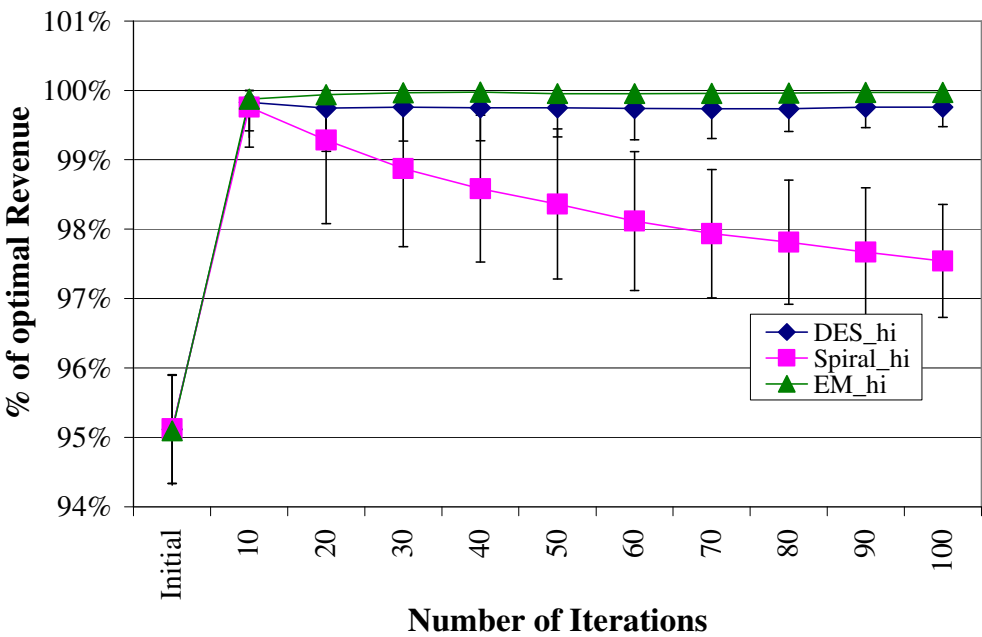
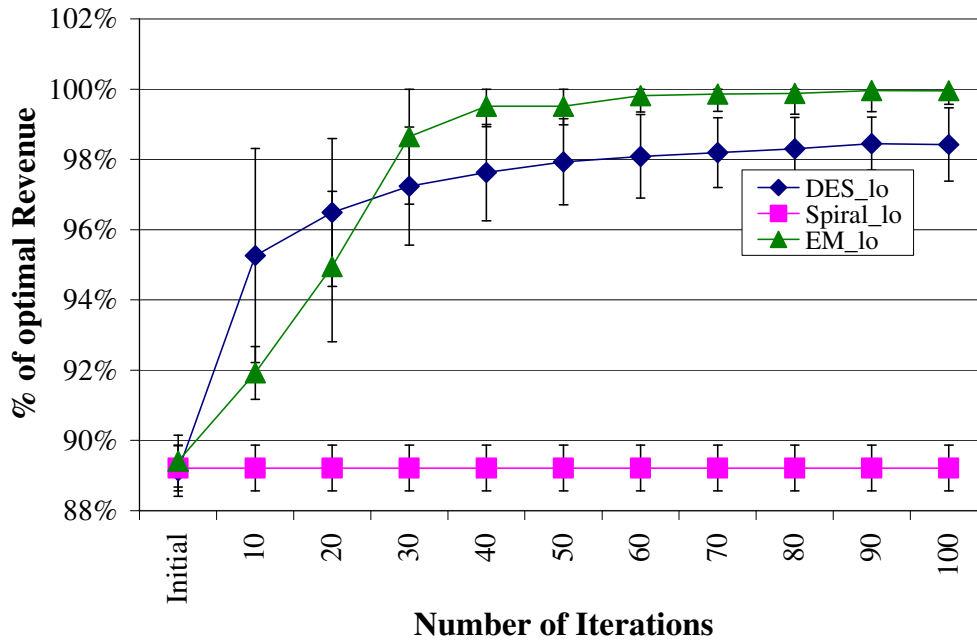


Figure 7 shows a convergence to the optimal revenue for DES and EM methods after starting with the high initial protection levels. Here both unconstraining methods converge to optimal protection levels and hence achieve optimal revenues. The DES and EM methods yield similar results, both starting at 95.5% of optimal revenues and improving to almost 100% after only one iteration. High starting protection levels restrict early bookings in the lower value segments while saving capacity for

the high value segments. When historical data is limited and the difference in revenue between high and low value segments is large, a firm may want to initially employ high protection levels.

Figure 8 - Revenue achieved using low protection limits for EM, DES, and no unconstraining



Compared to the high starting protection levels in Figure 7, the low starting protection levels in Figure 8 converge to the optimal revenue for both unconstraining methods at a much slower rate. Because all three of the highest fare classes are initially 100% constrained, and the EM method requires at least one unconstrained booking curve, we cannot use the EM method during the first iteration. Instead, we increase the protection limits by 10% for each group of 10 booking curves until at least one booking curve is not constrained, at which point we can begin using the EM method. While this practice may seem arbitrary (justifiably so), it is representative of techniques commonly applied in practice. The DES method does not suffer from such a limitation, thus it outperforms the EM method during the early stages of the low starting protection levels case. Once the EM method can be used, revenue quickly

converges to greater than 99% of optimal revenue. DES performs better in early iterations, then converges to just above 98% of optimal revenues, slightly trailing the EM method's performance.

Both Figures 7 and 8 illustrate that unconstrained data may lead to a loss in revenue. When starting with low protection levels, Figure 8 shows the “no unconstraining” (Spiral) data never improves past the initial 89% of optimal revenue. This compares unfavorably with the two unconstraining methods which steadily improve as more demand is observed. Figure 7 shows that when using high initial protection levels, failing to unconstrain data causes revenue to decrease every time protection limits are recalculated as the historical data becomes more and more constrained. This shows graphically the “spiral-down effect” (Homem-de-Mello et al. 2006) mentioned earlier.

Because the revenue converges to the optimal much faster in Figure 7 than in Figure 8, one may conjecture from a comparison of Figures 7 and 8 it is always better to start with high protection levels versus low. However, such a generalization is incorrect as this phenomenon is an artifact of this particular hotel's fare class structure. For this hotel, the difference in revenues between the highest fare class (\$100) and the next highest (\$62) is much larger than the difference between the two lowest fare classes (\$35 and \$25). If these differences had been reversed (say fare classes of \$100, \$90, \$63, and \$25), starting with low initial protection levels would converge to the optimal revenue much faster than starting with high protection levels.

We note a few additional observations from our bootstrapping results. First, when demand has sudden shifts or a large number of constrained days, all statistical methods become less accurate. In practice, we recommend qualitative adjustments to the demand data or the numerical forecast for these situations. Second, statistical unconstraining should be supplemented by a physical constrained count, where possible, to check the validity of unconstrained forecasts. For a hotel, this physical count could include reservation agents and bellhops keeping a manual tally of the number of people turned away. A

store selling physical goods might better promote “rain checks” of sold out items and keep track of how many people ask for the sold out good. Third, EM becomes more accurate for larger sample sizes, but often performs poorly for small sample sizes. In these cases, DES may be a better solution until more historical data is available. Fourth, a small firm with a limited IT budget may not have the resources to afford a sophisticated statistical program to run the EM method. Such a firm may find it more useful to use the straightforward DES method to estimate total demand. Lastly, if data is fully constrained, EM does not work, and an alternate method must be used. Either an alternative unconstraining method or a “rule of thumb” adjustment to the protection level is needed until unconstrained demand is observed.

6. Conclusions/ Recommendations

This paper examines the often overlooked but essential topic of unconstraining sales data. True demand distribution parameters are a critical ingredient to revenue management systems; unfortunately the data available is often constrained. Ignoring the constrained data problem results in significant reductions in revenue and observing demand after it exceeds capacity is often impractical, thus statistical unconstraining methods are often used to estimate the parameters of the demand distribution. We propose a new unconstraining method (DES) based on a common forecasting model that, unlike traditional statistical unconstraining methods, takes into account the point in time on the historical demand booking curves that demand was constrained. We find our proposed DES method and the expectation maximization (EM) method perform similarly well and outperform the alternatives. When little historical data is available or all demand sets are constrained, DES is a better choice than EM. Also, if sophisticated statistical software packages are not available, DES provides a better alternative than the averaging method.

We test the revenue impact of DES, EM, and “no unconstraining” on actual booking data from a hotel/casino. We show performance varies with the initial starting protection limits and a lack of

unconstraining leads to significant revenue losses. In our example, starting with high(low) initial protection levels converge to optimal limits more quickly(slowly). Low initial protection levels can lead to completely constrained classes, forcing firms using the EM method for unconstraining to use other methods until some unconstrained booking curves are observed. Both EM and DES take numerous iterations (5 and 3, respectively) to converge to within 2% of optimal with low initial protection levels. On the other hand, both EM and DES converge to within 0.3% of optimal after only one iteration when starting with high protection levels. These results show that both EM and DES are effective methods for unconstraining and provide significant improvements in total revenues over performing no unconstraining on the demand data at all.

As is true with all research, there are limitations to our work. In our section on revenue impact, we assume independent demand for a given customer value segment. That is, a customer associated with value segment 2 will not turn up as demand in segments 1 or 3, even if value segment 2 is closed. This is a very realistic assumption in a casino application, where customer gaming habits are not dependent upon the rate program they book under, but rather vice versa. However, in many other revenue management applications – particularly airlines - demand for a given segment often depends on the choice of segments available. The independence assumption is commonly used however in most booking limit algorithms. To incorporate consumer choice behavior such as buy-up or buy-down behavior, Talluri and van Ryzin (2004a) present a model that explicitly accounts for the probabilities that customers in a given fare class (value segment) will purchase from other fare classes if their preferred fare class is unavailable. To use this model, a firm needs to know the probabilities that customers in all classes will buy-up, or buy-down, probabilities that are rarely known in practice. Further development of such models along with unconstraining methods to accommodate them is a promising area of future research. Additionally, our DES method does not account for seasonality or

price promotions. Like most other forecasting methods, historical data should be decomposed into components of promotion effects, seasonality, and competitive effects before DES is applied. For further information on these adjustments, see Armstrong (2001). Despite these limitations, this work contributes needed analytical tools to an emerging research field.

References

- Altay, N. and Litteral, L. 2005. "A Comparative Study of Intermittent Demand Forecasting Models", working paper, Robins School of Business, University of Richmond.
- Armstrong, J. Scott (ed.). 2001. *Principles of Forecasting: A Handbook for Reserachers and Practitioners*. Norwell, MA: Kluwer Academic Publishers.
- Badinelli, R.D. 2000. "An Optimal, Dynamic Policy for Hotel Yield Management," *European Journal of Operational Research* **121**: 476-503.
- Belobaba, P.P. 1989. "Application of a Probabilistic Decision Model to Airline Seat Inventory Control," *Operations Research* **37** (2): 183-197.
- Bitran, G.R. and S. M. Gilbert. 1996. "Managing Hotel Reservations with Uncertain Arrivals," *Operations Research* **44**: 35-49.
- Bitran, G.R. and S. V. Mondschein. 1995. "An Application of Yield Management to the Hotel Industry Consdiering Multiple Day Stays," *Operations Research* **43**: 427-443.
- Carroll, B. and J. Siguaw. 2003. "The Evolution of Electronic Distribution: Effects on Hotels and Intermediaries," *Cornell Hotel and Restaurant Administration Quarterly* **44**(4) 38-50.
- Cox, D.R. 1972. "Regression Models and Life Tables (with discussion)," *Journal of the Royal Statistical Society*. **B34** (2): 187-220.
- Cox, D.R. and D. Oakes. 1984. *Analysis of Survival Data*. New York, Chapman & Hall.
- Cross, R. G. 1997. *Revenue Management*. New York, Broadway books.
- Croston, J.D. 1972. "Forecasting and Stock Control for Intermittent Demands," *Operational Research Quarterly* **23** (3): 289-303.
- Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*. **39** (1): 1-38.
- Homem-de-Mello, T., W.L. Cooper, A.J. Kleywegt. 2006. "Models of the Spiral-Down Effect in Revenue Management," To be published in *Operations Research*.
- Kalbfleisch, J.D. and R.L. Prentice. 1980. *The Statistical Analysis of Failure Time Data*. New York, John Wiley.
- Lawless, J.F. 1982. *Statistical Models and Methods for Lifetime Data*. New York, John Wiley.
- Liu, P.H. and V. Makis. 1996. "Cutting-tool reliability assessment in variable machining conditions," *IEEE Transactions on Reliability*. **45**(4): 573-581.

- Liu, P.H, S. Smith, E.B. Orkin, and G. Carey. 2002. "Estimating unconstrained hotel demand based on censored booking data," *Journal of Revenue and pricing Management*. **1**(2): 121-138.
- Maddala, G.S. 1983. *Limited Dependent And Qualitative Variables in Econometrics*. Cambridge, Cambridge University Press.
- McGill, J.I. 1995. "Censored regressed analysis of multiclass passenger demand data subject to joint capacity constraints," *Annals of Operations Research* **60**:209-240.
- McGill, J.I. and G.J. van Ryzin. 1999. "Revenue Management: Research Overview and Prospects," *Transportation Science* **33**(2): 233-256.
- Nelson, W. 1990. *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses*. New York, John Wiley.
- Orkin, E.B. 1998. "Wishful Thinking and Rocket Science," *Cornell Hotel and Restaurant Administration Quarterly* **39**(4) 15-20.
- Rothstein, M. 1974. "Hotel Overbooking as a Markovian Sequential Decision Process," *Decision Sciences* **5**:389-404.
- Schneider, H. 1986. *Truncated and Censored Samples from Normal Populations*. New York, Marcel Dekker.
- Shenstone, L. and R. J. Hyndman. 2005. "Stochastic Models Underlying Croston's Method for Intermittent Demand Forecasting," *Journal of Forecasting* **24** (6) :389-402.
- Talluri, K. and G.J. van Ryzin. 2004a. "Revenue Management Under a General Discrete Choice Model of Consumer Behavior," *Management Science* **50** (1): 15-33.
- Talluri, K. and G.J. van Ryzin. 2004b. *The Theory and Practice of Revenue Management*, Kluwer Academic Publishers, Norwell, Massachusetts.
- van Ryzin, G.J. and J.I. McGill. 2000. "Revenue Management Without Forecasting or Optimization: An Adaptive Algorithm for Determining Airline Seat Protection Levels," *Management Science* **50** (1): 15-33.
- Weatherford, L.R. and S.E. Bodily. 1992. "A Taxonomy and Research Overview of Perishable-Asset Revenue Management: Yield Management, Overbooking, and Pricing," *Operations Research* **40**(5):831-844.
- Weatherford, L.R. and S. Polt. 2002. "Better Unconstraining of Airline Demand Data in Revenue Management Systems for Improved Forecast Accuracy and Greater Revenues," *Journal of Revenue and Pricing Management* **1**(3):234-254.

Zeni, R. 2001. “Improving Forecast Accuracy by Unconstraining Censored Demand Data” AGIFORS, Bangkok, Thailand.

Appendix

Applying the DES method to booking curves that were opened and closed multiple times (Section 3)

Our DES method can be used when a booking curve is opened and closed multiple times. The methodology directly follows from the steps used when a booking curve is only closed once. To demonstrate, refer again to Figure 1, where the booking curve is closed after $t=30$. Imagine now that the booking class is reopened after $t=15$, then closed again after $t=5$ for the remaining five periods. In this example, we must estimate the demand occurring from $t=29$ to $t=15$, then again from $t=4$ to $t=0$.

For this example, our method works as follows. First, we estimate the unseen demand from day 29 – 15 (instead of 29 – 0 in our earlier example from Section 3). These smoothing coefficients are used to recalculate FIT_t ; call the new forecasts $\overline{FIT}_t = \overline{F}_t + \overline{T}_t$, for periods $t = 140 - 30$. We estimate the cumulative demand up to the closest period where demand is not constrained, \overline{FIT}_{15} in our example, using $\overline{FIT}_{15} = \overline{F}_{30} + 15\overline{T}_{30}$. Because we know the actual demands that occurred during days 14 to 5, we calculate \overline{FIT}_{14} to \overline{FIT}_5 using the recursive methods of (1), (2) and (3) for each of these days until we reach $\overline{FIT}_5 = \overline{F}_5 + \overline{T}_5$, after which we no longer know true demand and can not continue to update our forecast. As before, our objective is to determine the final cumulative demand (if demand was not constrained) at the terminating period $t = 0$. We estimate this value using $\overline{FIT}_0 = \overline{F}_5 + 5\overline{T}_5$.

Description of Unconstraining Methods (Section 4)

Averaging Method (AM) – As described by Weatherford and Polt (2002) – Replace ‘closed’ observations with the larger of (a) actual observation or (b) average of the ‘open’ observations.

Calculate the average and standard deviation parameter estimates based on open observations and modified values of closed observations.

The AM is demonstrated using the airline booking data example from Weatherford and Polt (2002). In this example, there are 10 observations for total cumulative demand from 10 different booking curves: 18, 19*, 17*, 19, 20, 17, 19*, 21*, 18, 20*. The * next to a number indicates an observation where demand reached the booking limit before departure, hence, constrained demand. The average of the open observations (18, 19, 20, 17, 18) is 18.4. Using the AM, we use the following data for unconstrained demand: 18, 19, 18.4, 19, 20, 17, 19, 21, 18, 20. In other words, unconstrained observations are left as is, and constrained observations are replaced by the larger of (a) the observed demand or (b) the average of the unconstrained observations. Hence, the closed observation of 17 is replaced by 18.4 and the closed observations of 19, 19, 21, and 20 are left as is.

Expectation Maximization (EM) – As explained in Talluri and van Ryzin (2004b) –

Suppose we have $M+N$ observations of bookings for a given product, z_1, \dots, z_{M+N} , of which M observations are constrained because the product was closed. We ignore the time-series aspect of the observations and treat z_1, \dots, z_{M+N} as an unordered set of observations generated by an i.i.d. process.... Our goal is to find the parameters of an underlying demand distribution for these observations.

Assume that the underlying demand distribution is normal with mean μ and standard deviation σ . We further assume that all the observations come from a common distribution and that the observations are constrained at random, i.e., they appear randomly in the sample. Since we are treating the observations as unordered, assume z_1, \dots, z_M are constrained at booking limits b_1, \dots, b_M , so that $z_1=b_1, \dots, z_M=b_M$. The remaining N observations are unconstrained.

If the data were not constrained, then it would be easy to construct the complete-data likelihood function. Namely,

$$L(\mu, \sigma, M + N) = \prod_{i=1}^{M+N} \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(z_i - \mu)^2}{2\sigma^2}}$$

with the complete-data log-likelihood function given by

$$\ln L(\mu, \sigma, M + N) = -\frac{M + N}{2} \ln 2\pi - (M + N) \ln \sigma - \frac{\sum_{i=1}^{M+N} (z_i - \mu)^2}{2\sigma^2}.$$

The μ and σ that maximizes $\ln L()$ in the above equation are given by the closed-form solution

$$\hat{\mu} = \frac{1}{M + N} \sum_{i=1}^{M+N} z_i$$

$$\hat{\sigma}^2 = \frac{1}{M + N} \sum_{i=1}^{M+N} (z_i - \hat{\mu})^2$$

However, we do not know the true values of the M censored observations z_1, \dots, z_M and therefore cannot use this procedure directly. Instead, we use the following steps:

Step 0 (Initialize): Initialize μ and σ to be $\mu^{(0)}$ and $\sigma^{(0)}$. Good candidates for these starting values are the sample mean and sample standard deviation of all the uncensored observations. Let $\delta > 0$ be a small number, to be used as a stopping criterion.

$$\mu^{(0)} = \frac{\sum_{i=M+1}^{M+N} z_i}{N}$$

$$\sigma^{(0)} = \sqrt{\frac{\sum_{i=M+1}^{M+N} (z_i - \mu^{(0)})^2}{N}}$$

Step 1 (E-step): Calculate the expected value of the censored data in the log-likelihood function

assuming that they come from a normal distribution X with parameters $(\mu^{(k-1)}, \sigma^{(k-1)})$. That is, for $i =$

1, ..., M calculate

$$\hat{Z}_i^{(k-1)} \doteq E[X \mid X \geq b_i, X \sim N(\mu^{(k-1)}, \sigma^{(k-1)})] \text{ and}$$

$$(\hat{Z}_i^2)^{(k-1)} \doteq E[X^2 \mid X \geq b_i, X \sim N(\mu^{(k-1)}, \sigma^{(k-1)})]$$

Next, for each censored observation $i = 1, \dots, M$, replace z_i by $\hat{Z}_i^{(k-1)}$ and z_i^2 by $(\hat{Z}_i^2)^{(k-1)}$ to form the complete-data log-likelihood function $Q(\mu, \sigma)$. In this way, we replace the constrained values in the log-likelihood function by their expected values given the current estimates of the mean and standard deviation.

Step 2 (M-step): Maximize $Q(\mu, \sigma)$ with respect to μ and σ to obtain $\mu^{(k)}$, $\sigma^{(k)}$, yielding

$$\mu^{(k)} = \frac{1}{M+N} \left[\sum_{i=1}^M \hat{Z}_i^{(k-1)} + \sum_{i=M+1}^{M+N} z_i \right] \text{ and}$$

$$\sigma^{(k)} = \frac{1}{M+N} \left[\sum_{i=M+1}^{M+N} \left((z_i^2)^{(k-1)} - 2\hat{Z}_i \mu^{(k-1)} + (\mu^{(k-1)})^2 \right) + \sum_{i=M+1}^{M+N} (z_i - \mu^{(k-1)})^2 \right]$$

Step 3 (Convergence test): If $\|\mu^{(k)} - \mu^{(k-1)}\| < \delta$ and $\|\sigma^{(k)} - \sigma^{(k-1)}\| < \delta$, then stop; else, $k \leftarrow k+1$, goto step 1.

Once convergence has been achieved – say, in iteration K – the unconstrained values for z_i , $i = 1, \dots, M$ can be taken as $E[X | X \geq b_i]$, where X is normally distributed with mean $\mu^{(K)}$ and standard deviation $\sigma^{(K)}$.

Projection Detruncation (PD) – as explained in Weatherford and Polt (2002) – This is a variation of the EM method, which uses a weighted value, τ , to scale the aggressiveness of unconstraining. Instead of calculating a conditional expectation, it iteratively finds a mean and standard deviation such that $\tau =$ area under normal curve from the unconstrained mean to infinity divided by the area under the normal curve from the constrained mean to infinity. The user chooses τ , where $0 \leq \tau \leq 1$. $\tau=0.5$ performs similarly to EM, using the conditional median rather than a conditional mean. $\tau \leq 0.5$ produces larger estimates for distribution parameters than $\tau \geq 0.5$.

Lifetables (LT) – as described by van Ryzin and McGill (2000) – Let n denote the total number of observations (censored and uncensored). Let $t_1 < t_2 < \dots, t_m$ be m distinct intervals. (Call $[t_j, t_{j+1})$

interval j .) Let n_j be the number of observations with values t_j or more (the number of “at risk” observations at the start of interval j); let d_j be the number of uncensored observations that fall in interval j (the number of “deaths” in the interval j); and let w_j be the number of censored observations that fall in interval j (the number of “withdrawals” because of censoring in interval j). Define $n_0 = n$ and note that $n_j = n_{j-1} - d_j - w_j, j = 1, \dots, m$.

The life table estimate is given by

$$\hat{S}\left(\frac{t_j + t_{j+1}}{2}\right) = \prod_{i=1}^j \left(1 - \frac{d_i}{n_i - w_i/2}\right)$$

Each term $1 - (d_i/(n_i - w_i/2))$ is an estimate of the conditional probability that demand exceeds t_{i+1} given that it exceeded t_i . The denominator, $n_i - w_i/2$, is an estimate of the number of samples at risk during period i , where $w_i/2$ is a correction term for the number of censored observations in period i . We used 20 intervals in our work.

The life table estimator can now be used to estimate the mean and standard deviation of the distribution by linear regression. Specifically, let $\Phi(x)$ be the standard normal distribution and let $\Phi^{-1}(x)$ denote its inverse. Define $s_j = \Phi^{-1}(1 - \hat{S}((t_j + t_{j+1})/2))$. If demand is normally distributed, the points $(s_j, t_j), j = 1, \dots, n$ should lie approximately on a straight line, namely $s_j = at_j + b$. Using linear regression, we use our estimates of the slope, \hat{a} , and the intercept, \hat{b} , to estimate the mean, $\hat{\mu} = 1/\hat{a}$, and the standard deviation, $\hat{\sigma} = -\hat{b}/\hat{a}$.

Creation of sample booking curves (Section 4)

In this section we explain how we created the three booking curves (concave, convex, and homogeneous) used to test the unconstraining methods in Section 4.

Concave

We created 100 concave booking curves which simulates arrival rates over a 140 day period. In order to create a concave booking curve, the arrival rate must decrease over the 140 day booking period. Hence, for days 140 – 121, we randomly generated arrival rates for each day out of a Poisson distribution with $\lambda=8$. Next, we randomly generated arrival rates for days 120-101 from a Poisson distribution with $\lambda=7$. We continued in this manner, decreasing λ by 1 every 20 days until we reached days 20 – 1 with $\lambda=2$. After simulating one booking curve in this manner, we repeated this procedure 99 times in order to generate 100 concave booking curves.

These 100 concave booking curves have an average total demand, $\mu = 700.20$, and a standard deviation, $\sigma = 26.96$. Using the previously stated formula for booking limits, $\text{Booking limit} = \mu + z^* \sigma$, we found a target booking limit in order to constrain 20%, 40%, 60%, 80%, and 98% of all booking curves (values found in table below).

% of booking curves constrained	20%	40%	60%	80%	98%
Target booking limit	722.89	707.03	693.37	677.51	644.83

Given the 100 concave booking curves and given the target booking limit of 722.89 for constraining 20% of the total booking curves, we generated random booking limits for each of 100 booking curves. We randomly generated 100 numbers out of the normal distribution with $\mu=722.89$ and $\sigma=26.96$. Those 100 numbers served as the booking limit for the 100 booking curves. Hence, if the first curve had a total demand of 726 and its corresponding booking limit was 720, the first curve had constrained demand. Once the booking limits were matched to the booking curves, we determined which curves were constrained or unconstrained. Next, we unconstrained each constrained booking curve using each of the described methods (AM, EM, PD, DES, LT). We then calculated the average and standard deviation of total demand across our unconstrained data and compared this value to the average and

standard deviation of total demand across the true demand. This simulation was repeated for the scenarios where we constrained 40%, 60%, 80%, and 98% of all booking curves.

Convex

The convex booking curves were created similarly to the concave curves. In order to create a convex booking curve, the arrival rate must increase over the 140 day booking period. Hence, for days 140 – 121, we randomly generated arrival rates for each day out of a Poisson distribution with $\lambda = 2$. Next, we randomly generated arrival rates for days 120-101 from a Poisson distribution with $\lambda = 3$. We continued to increase λ by 1 after 20 days of arrival until we reached days 20 – 1 with $\lambda = 8$. After simulating one booking curve in this manner, we repeated it 99 times in order to generate 100 convex booking curves.

The 100 convex booking curves had an average total demand, $\mu = 696.74$, and standard deviation, $\sigma = 27.93$. Just as with the concave data set, we used the booking limit formula to find target booking limits in order to constrain 20%, 40%, 60%, 80%, and 98% of all booking curves.

% of booking curves constrained	20%	40%	60%	80%	98%
Target booking limit	720.25	703.82	689.66	673.23	639.38

For the 20% constrained level, we randomly generated 100 numbers out of the normal distribution with $\mu = 720.25$ and $\sigma = 27.93$. Those 100 numbers served as the booking limit for the 100 booking curves. Hence, if the first curve had a total demand of 726 and its corresponding booking limit was 720, the first curve had constrained demand. Once the booking limits were matched to the booking curves, we determined which curves were constrained or unconstrained. Next, we unconstrained each constrained booking curve using each of the described methods (AM, EM, PD, DES, LT). We then calculated the average and standard deviation of total demand across our unconstrained data and

compared this value to the average and standard deviation of total demand across the true demand. This simulation was repeated for the scenarios where we constrained 40%, 60%, 80%, and 98% of all booking curves.

Homogeneous

We created 100 homogeneous booking curves by randomly generating 140 arrival rates out of a Poisson distribution with $\lambda = 5$. After simulating one booking curve in this manner, we repeated it 99 times in order to generate 100 homogeneous booking curves. The 100 booking curves average had an average total demand = 698.14 and a $\sigma = 26.52$. Using the booking limit formula, we found a target booking limit in order to constrain 20%, 40%, 60%, 80%, and 98% of all booking curves.

% of booking curves constrained	20%	40%	60%	80%	98%
Target booking limit	720.46	704.86	691.42	675.82	643.67

Booking limits were formed and tested as in the convex and concave cases.

DES for small and intermittent demand (Section 4.2)

In this section we provide a review of Croston's Method as described by Shenstone and Hyndman (2005). Let Y_t be the demand occurring in time period t and X_t be the indicator variable for non-zero demand periods; i.e., $X_t = 1$ when demand occurs at time period t and $X_t = 0$ when no demand occurs. Furthermore, let j_t be the index of the non-zero demand such that $j_t = \sum_{i=1}^t X_i$. Let Y_j^* represent the size of the j th non-zero demand and Q_j the inter-arrival time between Y_{j-1}^* and Y_j^* .

Croston's method separately forecasts the time between non-zero demand and the size of non-zero demand using simple exponential smoothing. Let Z_j and P_j be the forecasts of the $(j+1)$ th demand size and inter-arrival time respectively, based on data up to demand j . Croston's method gives

$$Z_j = (1-\alpha) Z_{j-1} + \alpha Y_j^* \text{ and}$$

$$P_j = (1-\alpha) P_{j-1} + \alpha Q_j .$$

The smoothing parameter α ranges from 0 to 1 and is chosen as the value that minimizes actual demand and inter-arrival time (respectively) from projected inter-arrival time at time $= j$. We assume α is the same for both Y_j^* and Q_j . Let $l = j_n$ denote the last period of demand. Then the mean demand rate, which is used for the next non-zero demand at time $n + h$, is estimated by the ratio $Y_{n+h} = Z_l / P_l$.

Example calculation of total revenue for a given demand realization (Section 5.3)

To demonstrate our method for calculating the percent of optimal revenue obtained by using inaccurate demand distribution parameters, we refer back to the example data given in Table 4. Suppose a particular unconstraining method gives the incorrect parameter estimates for fare class 1 of a mean and variance of 45 (as opposed to the correct parameter values of 50 for each). If all the other parameter estimates were the same, using the incorrect parameter estimates in the EMSR-b heuristic would result in nested protection levels of 44, 120, and 252 for fare classes 1, 2, and 3 respectively. Now suppose a particular realization for all four class booking curves results in total demand for fare classes 1, 2, 3, and 4 of 51, 75, 135, and 510 respectively. Assume the capacity of the hotel is 500 rooms. The total revenue using the protection levels calculated from the true distribution parameter values (Table 4) is detailed in the following calculations:

$$I_4 = Capacity - \theta_3 = 500 - 257 = 243$$

$$I_3 = Capacity - \theta_2 - \sum_{j=4}^4 \min(I_j, X_j) = 500 - 125 - \min(243, 510) = 132$$

$$I_2 = Capacity - \theta_1 - \sum_{j=3}^4 \min(I_j, X_j) = 500 - 49 - [\min(132, 135) + \min(243, 510)] = 76$$

$$I_1 = Capacity - \sum_{j=2}^4 \min(I_j, X_j) = 500 - [\min(76, 75) + \min(132, 135) + \min(243, 510)] = 50$$

$$R_i = f_i^* \min [X_i, I_i]$$

$$R_1 = f_1^* \min [X_1, I_1] = 250 * 50 = 12500$$

$$R_2 = f_2^* \min [X_2, I_2] = 150 * 75 = 11250$$

$$R_3 = f_3^* \min [X_3, I_3] = 100 * 132 = 13200$$

$$R_4 = f_4^* \min [X_4, I_4] = 50 * 243 = 12150$$

$$\text{Total Revenue} = \sum_{i=1}^n R_i = \$49,100.$$

Using equations 6, 7, and 8, but assuming incorrect protection levels of 44, 120, and 252 yields total revenue of \$48,100. Thus, using the incorrect protection levels from this example results in total revenues that are 98% (48100/49100) of optimal for this particular demand realization. Although every realization of demand does not result in higher total revenue for the correct protection levels versus the incorrect protection levels, on average the correct protection levels yield higher total revenues than the incorrect protection levels.